Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 1, 566–579
https://doi.org/10.1214/19-AIHP973

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Bounds on the Poincaré constant for convolution measures

Thomas A. Courtade

University of California, Berkeley, Department of Electrical Engineering and Computer Sciences, 265 Cory Hall, Berkeley CA 94720-1770, USA.
E-mail: courtade@berkeley.edu

Received 21 September 2018; revised 14 January 2019; accepted 13 February 2019

Abstract. We establish a Shearer-type inequality for the Poincaré constant, showing that the Poincaré constant corresponding to the convolution of a collection of measures can be nontrivially controlled by the Poincaré constants corresponding to convolutions of subsets of measures. This implies, for example, that the sequence of Poincaré constants corresponding to successive convolutions in the central limit theorem is non-increasing. We also establish a dimension-free stability estimate for subadditivity of the Poincaré constant on convolutions which uniformly improves an earlier one-dimensional estimate of a similar nature by Johnson (*Teor. Veroyatn. Primen.* **48** (2003) 615–620). As a byproduct of our arguments, we find that the various monotone properties of entropy, Fisher information and the Poincaré constant on convolutions have a common, simple root in Shearer's inequality.

Résumé. Nous démontrons une inégalité de type Shearer pour les constantes de Poincaré, selon laquelle la constante correspondant à la convolution d'une famille de mesures peut être contrôlée de manière non-triviale par celles de convolutions de sous-familles. Ceci implique, par exemple, que les constantes de Poincaré décroissent de manière monotone le long du théorème central limite. Nous démontrons également une estimée de stabilité indépendante de la dimension pour la sous additivité des constantes de Poincaré de convolutions, améliorant un résultat unidimensionnel similaire dû à Johnson (*Teor. Veroyatn. Primen.* **48** (2003) 615–620). Comme conséquence de nos arguments, nous montrons que les diverses propriétés de monotonie de l'entropie, de l'information de Fisher et de la constantes de Poincaré pour les convolutions trouvent une même source en l'inégalité de Shearer.

MSC: 60E15; 39B62; 26D10

Keywords: Functional inequalities; Poincaré inequalities; Stability; Convolution measures

1. Introduction

Let $\mathcal{P}(\mathbb{R}^d)$ denote the set of Borel probability measures on $\mathbb{R}^d$. A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is said to satisfy a Poincaré inequality with constant C if

$$\mathrm{Var}_\mu(f) \le C \int_{\mathbb{R}^d} |\nabla f|^2 \, d\mu \tag{1.1}$$

for all locally Lipschitz functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, where ∇ denotes the usual gradient and $|\cdot|$ denotes the Euclidean length on $\mathbb{R}^d$. The Poincaré constant $C_p(\mu)$ is defined to be the smallest constant C for which (1.1) holds.

Poincaré inequalities play a central role in concentration of measure (see, e.g., [20, Ch. 3]), and imply dimension-free concentration inequalities for the product measures μ^n, $n \ge 1$, which depend only on the Poincaré constant $C_p(\mu)$. Indeed, it is an easy exercise to see that $C_p(\mu^n) = C_p(\mu)$, so the Poincaré inequality directly implies

$$\mathrm{Var}_{\mu^n}(f) \le C_p(\mu) \|f\|_{\mathrm{Lip}}^2 \quad \forall f : \mathbb{R}^{nd} \longrightarrow \mathbb{R},$$

where $\|\cdot\|_{\mathrm{Lip}}$ is the usual Lipschitz seminorm. Stronger concentration estimates are also available. For example, Bobkov and Ledoux [5] established the following dimension-free estimate for exponential concentration

$$\mu^n\left(f \geq \int f\,d\mu^n + t\right) \leq \exp\left(-\min\left(\frac{t^2}{KC_p(\mu)}, \frac{t}{\sqrt{KC_p(\mu)}}\right)\right), \tag{1.2}$$

holding for all 1-Lipschitz f, where K is an absolute constant. The fact that a Poincaré inequality implies dimension-free exponential concentration for Lipschitz functions has a long history, dating back to work by Gromov and Milman [16]. A converse statement also holds, implying that dimension-free concentration is equivalent to existence of a Poincaré inequality in a precise sense [15]. Thus, any information about $C_p(\mu)$ reveals quantitative information about the concentration properties enjoyed by μ. Beyond concentration of measure, Poincaré inequalities play an important role throughout analysis, for example in analysis of PDEs (e.g., [12]), and in characterizing convergence rates of stochastic dynamics (e.g., [2, Chapter 4] and references therein).

Except in special cases, $C_p(\mu)$ is not known explicitly for general probability measures μ, but it can sometimes be controlled using properties enjoyed by the Poincaré constant. For example, it is easy to check by change of variables that $C_p(\mu_{\alpha,\beta}) = \alpha^2 C_p(\mu)$, where $\mu_{\alpha,\beta}$ denotes the law of $\alpha X + \beta$, with $\alpha, \beta \in \mathbb{R}$ and $X \sim \mu$. Only slightly less immediate is the subadditivity property

$$C_p(\mu \star \nu) \leq C_p(\mu) + C_p(\nu) \tag{1.3}$$

for the convolution measure $\mu \star \nu$, which follows by a classical variance decomposition and convexity of $t \mapsto t^2$ (e.g., [6]).

It is convolution inequalities like (1.3) that are the focus of this paper. There have been several recent results along these lines which we now mention. For example, Bardet, Gozlan, Malrieu and Zitt [3] recently established dimension-free bounds on the Poincaré constant for Gaussian convolutions of compactly supported measures. Johnson [19] had obtained similar bounds on the Poincaré constant for finite mixtures of one-dimensional Gaussians with identical variances, and he further studied convergence of the Poincaré constant in the central limit theorem. This latter topic is closely related to some of the results contained in this paper, so we highlight the similarities and differences in the relevant sections. In a related direction, Chafai and Malrieu [7] gave bounds on the Poincaré constant for two-point mixtures. We remark that other bounds are known for the logarithmic Sobolev constant for convolution measures (which immediately yield bounds on the Poincaré constant), e.g. [3,26], but these tend to be weaker than estimates which target the Poincaré constant directly.

Our results fall into two main categories. First, we establish a Shearer-type inequality for the Poincaré constant, which shows that the Poincaré constant corresponding to the convolution of a collection of measures can be nontrivially controlled by the Poincaré constants corresponding to convolutions of subsets of measures. This has new and interesting consequences; for example, the Poincaré constant is monotone on the sequence of convolutions in the central limit theorem, similar to entropy. Second, we establish a dimension-free quantitative stability estimate for the inequality (1.3), which depends on the measures μ, ν only through their Poincaré constants. This uniformly improves upon a previous estimate of Johnson [19], which required a bound on Fisher information. The proofs all rely on a particular variance inequality, closely related to Shearer's lemma, which may be of independent interest. As a byproduct of our arguments, we see that monotone properties of entropy, Fisher information and the Poincaré constant find a common root in Shearer's inequality.

2. Presentation of results

2.1. *Bounds on the Poincaré constant for convolution measures*

Our first result is the following bound on the Poincaré constant for convolutions:

Theorem 2.1. *Let $(\mu_i)_{1\leq i\leq n} \subset \mathcal{P}(\mathbb{R}^d)$. For a set $S \subset [n] := \{1, 2, \dots, n\}$, let μ_S denote the convolution of $(\mu_i)_{i\in S}$. If $\mathcal{C}$ is a collection of distinct subsets of $[n]$, then*

$$C_p(\mu_{[n]}) \leq \frac{1}{t}\sum_{S\in\mathcal{C}} C_p(\mu_S) \tag{2.1}$$

where $t := \min_{i\in[n]} \#\{S \in \mathcal{C} : S \ni i\}$.

As an example, if we take $n = 2$ and $\mathcal{C} = \{\{1\}, \{2\}\}$, we see that (2.1) extends the classical subadditivity estimate (1.3). Two further examples in the case of convolutions of identical measures – where the expressions are simplest, but the results still new and illustrative – are given below:

Example 2.1. Let (X_i) be i.i.d. random vectors in $\mathbb{R}^d$ with law ν_1. For $n \geq 1$, let ν_n denote the law of the standardized sum $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. By taking $\mu_i = \nu_1$ for $1 \leq i \leq n$ and $\mathcal{C} = \{\{[n] \setminus i\} : i \in [n]\}$ in Theorem 2.1, we find

$$C_p(\nu_n) \leq C_p(\nu_{n-1}). \tag{2.2}$$

That is, the Poincaré constant is monotone on the sequence of convolutions in the central limit theorem.

Like entropy and Fisher information, monotonicity of $C_p(\nu_n)$ is suggested by its corresponding subadditivity property (though, not implied by it). The connection between (2.2) and monotonicity of entropy and Fisher information runs deep, and is articulated in Section 3.1 (Remark 3.2).

The following example is more counterintuitive, and does not seem to be predicted by subadditivity:

Example 2.2. Let ν_n be as in the previous example, and let γ_{δ^2} denote the law of the normal distribution $N(0, \delta^2 \mathrm{I})$. Applying Theorem 2.1 with $\mu_i = \nu_1$ for $1 \leq i \leq n$, $\mu_{n+1} = \gamma_{\delta^2}$, and $\mathcal{C} = \{\{i, n+1\} : i \in [n]\}$, we find

$$C_p(\nu_n \star \gamma_{\delta^2/n}) \leq C_p(\nu_1 \star \gamma_{\delta^2}), \quad \text{for each } n \geq 1. \tag{2.3}$$

The surprise here is that the degree of Gaussian regularization on the left is significantly less than that on the right (i.e., variance δ^2/n instead of δ^2), but the Poincaré constant is no worse.

To add to Example 2.2, we remark that discrete probability distributions are typical examples of measures that do not satisfy a Poincaré inequality, since one can always find a non-constant function f which remains constant on the support, ensuring $\nabla f = 0$ on sets of positive measure. To give a potentially counterintuitive example, illustrated in Figure 1, we can take ν_1 in (2.3) to be the equiprobable measure on $\{-\frac{1}{2}, \frac{1}{2}\}$, and $\delta^2 \ll 1$. In this case, the measure $\nu_n \star \gamma_{\delta^2/n}$ looks nearly discrete as n becomes fairly large (i.e., like a standardized Binomial$(n, 1/2)$ distribution), yet satisfies a Poincaré inequality with constant depending only on δ^2, e.g., $C_p(\nu_n \star \gamma_{\delta^2/n}) \leq C_p(\nu_1 \star \gamma_{\delta^2}) \leq \delta^2 \exp(4/\delta^2)$. To contrast this concrete example with previous estimates, an application of [3, Theorem 1.2] gives $C_p(\nu_n \star \gamma_{\delta^2/n}) \leq \frac{\delta^2}{n} \exp(4n^2/\delta^2)$, and Johnson's estimate for univariate Gaussian mixtures [19, Theorem 1.4] gives $C_p(\nu_n \star \gamma_{\delta^2/n}) \leq 2^n \exp(n 2^n/\delta^2)$. These bounds may be improved slightly using subadditivity, but nevertheless remain exponential in n.

2.1.1. *Application to a quantitative CLT in W_2*

The ability to introduce vanishing regularization via (2.3) may be useful in applications. One nice illustration is the following dimension-free quantitative central limit theorem in the L^2-Wasserstein distance on $\mathcal{P}(\mathbb{R}^d)$, denoted by W_2:

Corollary 2.1. *Let (X_i) be i.i.d. centered isotropic random vectors in $\mathbb{R}^d$ with law ν_1, and let ν_n denote the law of $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$. For any $\delta \geq 0$,*

$$W_2(\nu_n, \gamma_1)^2 \leq d \frac{4(\delta^2 + C_p(\nu_1 \star \gamma_{\delta^2}))}{\delta^2 + \sqrt{n} - 1}. \tag{2.4}$$

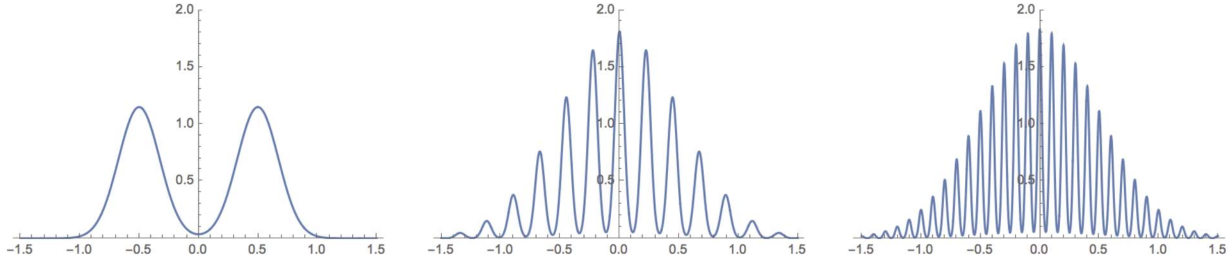

Fig. 1. From left to right, three probability densities with progressively better (i.e., non-increasing) Poincaré constants.

Proof. It suffices to assume that δ is such that $C_p(\nu_1 \star \gamma_{\delta^2}) < \infty$. The proof involves properties of the so-called Stein discrepancy $S(\mu|\gamma_1)^2$, which is a measure of the distance from a probability measure μ to the standard Gaussian measure γ_1. Its precise definition is not needed here, but we will need two properties. First, for any centered probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, finiteness of $C_p(\mu)$ implies finiteness of $S(\mu|\gamma_1)^2$ (see [9]). Combined with results from [21], if μ is centered and isotropic, then

$$W_2(\mu,\gamma_1)^2 \le S(\mu|\gamma_1)^2 \le \big(C_p(\mu)-1\big)d$$

and, with the notation ν_n prevailing,

$$S(\nu_{n+1}|\gamma_1)^2 \le S(\nu_n|\gamma_1)^2 \le \frac{1}{n} S(\nu_1|\gamma_1)^2 \quad \text{for all } n \ge 1.$$

For $n \ge 1$, let ν_n^t denote the law of $\frac{n^{-1/2}}{\sqrt{1+t}} \sum_{i=1}^n X_i$, so that $\nu_n^t \star \gamma_{t/(1+t)}$ is isotropic. Starting with the triangle inequality for W_2 and using each of the above estimates followed by (2.3), we have for any $t > 0$ and integers $n_1, n_2 \ge 0$ such that $n_1 n_2 \le n$,

$$\begin{aligned}
\frac{1}{2} W_2(\nu_n,\gamma_1)^2 &\le W_2\big(\nu_n^t \star \gamma_{t/(1+t)}, \nu_n\big)^2 + W_2\big(\nu_n^t \star \gamma_{t/(1+t)}, \gamma_1\big)^2 \\
&\le d\frac{2t}{1+t} + S\big(\nu_n^t \star \gamma_{t/(1+t)}|\gamma_1\big)^2 \\
&\le d\frac{2t}{1+t} + \frac{1}{n_1} S\big(\nu_{n_2}^t \star \gamma_{t/(1+t)}|\gamma_1\big)^2 \\
&\le d\frac{2t}{1+t} + \frac{d}{n_1} C_p\big(\nu_{n_2}^t \star \gamma_{t/(1+t)}\big) \\
&= d\frac{2t}{1+t} + \frac{d}{n_1}\left(\frac{1}{1+t} C_p(\nu_{n_2} \star \gamma_t)\right) \\
&\le d\frac{2t}{1+t} + \frac{d}{n_1}\left(\frac{1}{1+t} C_p(\nu_1 \star \gamma_{n_2 t})\right).
\end{aligned}$$

Now, choosing $t = \delta^2/n_2$ and $n_1 = n_2 = \lfloor\sqrt{n}\rfloor \ge \sqrt{n} - 1$, we simplify to find (2.4). □

A few remarks are in order: We emphasize that for (2.4) to apply nontrivially, ν_1 does not need to satisfy a Poincaré inequality, but only needs to have finite Poincaré constant after convolution with a Gaussian of sufficiently large variance.[1] This is a significantly weaker assumption than finiteness of the Poincaré constant; for instance, a simple modification of the proof of [3, Theorem 1.2] establishes that all sub-Gaussian distributions enjoy this property. Since $C_p(\mu^n) = C_p(\mu)$, the estimate (2.4) is dimension-free, and in fact has optimal dependence on dimension since W_2^2 is additive on product measures. However, the rate $O(n^{-1/2})$ is suboptimal, and should be $O(1/n)$ under moment constraints. In particular, Talagrand's inequality together with results of Bobkov, Chistyakov and Götze [4] imply an asymptotic rate of $W_2(\nu_n,\gamma_1)^2 = O(1/n)$ under finite fourth moment, but these estimates are non-quantitative in dimension greater than one, and prefactors generally behave poorly in dimension.

In [25], Zhai improved an earlier result by Valiant and Valiant [24] and established that if ν_1 is supported in the Euclidean ball of radius R, then it holds that

$$W_2(\nu_n,\gamma_1)^2 \le \frac{25dR^2(1+\log n)^2}{n}. \tag{2.5}$$

Since any isotropic measure supported in the Euclidean ball of radius R necessarily has $R^2 \ge d$, Zhai's estimate gives at best $O(d^2)$ scaling in the upper bound on W_2^2, which is worse than (2.4). However, (2.5) does offer the improved rate of $O((\log n)^2/n)$ compared to the rate of $O(n^{-1/2})$ in (2.4). As a result, if one is working with compactly supported distributions, then (2.4) would be preferred in the sample-limited regime where $n/(\log n)^4 \lesssim d^2$, and (2.5) would be preferred in the large-sample regime where $n/(\log n)^4 \gtrsim d^2$.

[1] In fact, if ν_1 has $C_p(\nu_1) < \infty$, then we can take $\delta = 0$, $n_1 = n$ and $n_2 = 1$ in the the proof of Corollary 2.1 to conclude $W_2(\nu_n,\gamma_1)^2 = O(dC_p(\nu_1)/n)$, which can be found in [9].

We also remark that near-optimal rates have very recently been obtained for the multivariate CLT in the weaker 1-Wasserstein distance under the assumption of finite third moments, albeit with suboptimal dependence on dimension [13].

2.1.2. *Remark on non-Euclidean settings*
Poincaré inequalities continue to make sense in settings beyond $\mathbb{R}^d$, but the applications remain similar (see, e.g., [14]). For example, if $(\mathcal{X}, d)$ is a metric space equipped with a probability measure μ, it is common to say μ satisfies a Poincaré inequality with constant C if

$$\mathrm{Var}_\mu(f) \le C \int_{\mathcal{X}} |\nabla f|^2 \, d\mu \tag{2.6}$$

for a sufficiently large class of test functions $f : \mathcal{X} \longrightarrow \mathbb{R}$. Here, the length of the gradient is defined as

$$|\nabla f|(x) := \limsup_{y \longrightarrow x} \frac{|f(x) - f(y)|}{d(x, y)} \tag{2.7}$$

whenever x is an accumulation point (otherwise $|\nabla f|(x) = 0$). Under this definition, the dimension-free concentration estimate (1.2) continues to hold [5].

It turns out that Theorem 2.1 can be extended to cover these and other situations. Let $(\mathcal{X}, +)$ be an Abelian group, where $\mathcal{X}$ is a Polish space, and let $\mathbb{B}(\mathcal{X}, \mathbb{R})$ denote the collection of bounded real-valued functions on $\mathcal{X}$. Consider a collection of functions $\mathcal{A} \subset \mathbb{B}(\mathcal{X}, \mathbb{R})$ which is closed under translation; i.e., $f \in \mathcal{A} \Leftrightarrow f(\cdot + t) \in \mathcal{A}$ for all $t \in \mathcal{X}$. Further, let ∇ be an operator on the elements of $\mathcal{A}$ which commutes with translation in the sense that $|\nabla(f(\cdot + t))| = |(\nabla f)(\cdot + t)|$ for all $f \in \mathcal{A}$ and $t \in \mathcal{X}$. In words, the length of the gradient of the map $x \mapsto f(x + t)$ is equal to the length of the gradient of f, evaluated at $x + t$. Note that this condition is met for (2.7), assuming the metric d is translation invariant. It is also satisfied by discrete derivatives in symmetric settings (e.g., the hypercube).

With the above definitions, for a given probability measure $\mu \in \mathcal{P}(\mathcal{X})$, define $C_p(\mu; \mathcal{A})$ to be the smallest constant C such that (2.6) holds for all $f \in \mathcal{A}$.

Theorem 2.2. *Let the above notation prevail, and consider a collection of probability measures $(\mu_i) \subset \mathcal{P}(\mathcal{X})$. For $S \subset [n]$, let μ_S denote the law of $\sum_{i \in S} X_i$, where $X_i \sim \mu_i$ are independent and summation is with respect to the group operation $+$. If $\mathcal{C}$ is a collection of distinct subsets of $[n]$, then*

$$C_p(\mu_{[n]}; \mathcal{A}) \le \frac{1}{t} \sum_{S \in \mathcal{C}} C_p(\mu_S; \mathcal{A})$$

where $t := \min_{i \in [n]} \#\{S \in \mathcal{C} : S \ni i\}$.

In applications, $\mathcal{A}$ will generally be dense in the class of test functions with respect to an appropriate norm. For example, in the case of $\mathcal{X} = \mathbb{R}^d$ where ∇ is the usual (weak) gradient, then Theorem 2.1 follows from Theorem 2.2 by density of smooth functions in the Sobolev space $W^{1,2}(\mathbb{R}^d, \mu)$, where μ has density with respect to Lebesgue measure.

2.2. *Stability of subadditivity of the Poincaré constant*

For $\mu \in \mathcal{P}(\mathbb{R}^d)$, define

$$\sigma^2(\mu) := \max_{\alpha \in \mathbb{R}^d : |\alpha| = 1} \mathrm{Var}_\mu(x \mapsto \alpha \cdot x)$$

to be the largest variance of μ in any direction (equivalently, the largest eigenvalue of the covariance matrix for μ). It is known that $C_p(\mu) \ge \sigma^2(\mu)$, with equality only if μ is marginally Gaussian in the direction of largest variance. In fact, as shown in [9], if $\alpha^* \in \arg\max_{\alpha \in \mathbb{R}^d : |\alpha| = 1} \mathrm{Var}_\mu(x \mapsto \alpha \cdot x)$, then

$$C_p(\mu) - \sigma^2(\mu) \ge W_2\big((\alpha^* \cdot \mathrm{id})\#\mu, \gamma_{\sigma^2(\mu)}\big)^2, \tag{2.8}$$

where γ_{σ^2} is the law of $N(0, \sigma^2(\mu))$, and $(\alpha^* \cdot \mathrm{id})\#\mu$ is the pushforward of μ under the map $x \mapsto \alpha^* \cdot x$ (i.e., $(\alpha^* \cdot \mathrm{id})\#\mu$ is the marginal distribution of μ in direction α^*). We remark that that the one-dimensional nature of (2.8) is unavoidable, since the Poincaré constant of μ is at least as bad as any one-dimensional marginal (e.g., consider product measures of the form $\gamma_{\sigma^2} \otimes \mu^n$).

Our main result of this section is a dimension-free stability estimate for the subadditivity property (1.3) in terms of the gap $C_p(\mu \star \nu) - \sigma^2(\mu \star \nu)$ (and, by (2.8), in terms of the non-gaussianness of any one-dimensional marginal of μ with largest variance). In particular,

Theorem 2.3. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, and define $\sigma^2 = \sigma^2(\mu \star \nu)$ for convenience. Then,*

$$C_p(\mu \star \nu) \le \big(C_p(\mu) + C_p(\nu)\big) - \frac{C_p(\mu)C_p(\nu)}{C_p(\mu) + C_p(\nu)} \frac{(C_p(\mu \star \nu) - \sigma^2)^2}{(C_p(\mu \star \nu) - \sigma^2)^2 + C_p(\mu \star \nu)\sigma^2}.$$

Of note, the above estimate is dimension-free, and requires no quantitative information about the measures beyond their Poincaré constants and the value $\sigma^2(\mu \star \nu)$. In order for equality to hold in (1.3), we need that $C_p(\mu \star \nu) = \sigma^2(\mu \star \nu)$, implying that $\mu \star \nu$ is marginally Gaussian in its direction of maximum variance by (2.8). Cramer's theorem then implies that μ and ν must be marginally Gaussian in this same direction as well.

Letting $\mu = \nu$ in Theorem 2.3, the following stability estimate for i.i.d. sums is immediate:

Corollary 2.2. *Let X_1, X_2 be i.i.d. random vectors in $\mathbb{R}^d$ with law ν_1, and define ν_2 to be the law of the standardized sum $\frac{1}{\sqrt{2}}(X_1 + X_2)$. Then*

$$C_p(\nu_2) \le C_p(\nu_1) - \frac{C_p(\nu_1)}{4} \frac{(C_p(\nu_2) - \sigma^2)^2}{(C_p(\nu_2) - \sigma^2)^2 + C_p(\nu_2)\sigma^2}, \tag{2.9}$$

where $\sigma^2 = \sigma^2(\nu_2) = \sigma^2(\nu_1)$.

Johnson's paper contains a result similar to Corollary 2.2, so we describe notable differences below. First, we mention that our results hold for probability measures on $\mathbb{R}^d$, whereas Johnson's results are derived only for dimension 1. Perhaps more substantially, the stability estimate derived by Johnson depends on Fisher information, whereas ours does not. Assuming $d = \sigma^2(\nu_1) = 1$, Johnson's key result can be stated in a form comparable to (2.9) as[2]

$$C_p(\nu_2) \le C_p(\nu_1) - \frac{C_p(\nu_1)}{9} \frac{(C_p(\nu_2) - 1)^2}{C_p(\nu_2)^2(1 + J(\nu_1)C_p(\nu_1))}, \tag{2.10}$$

where

$$J(\mu) := \int_{\mathbb{R}} \frac{|f'(x)|^2}{f(x)}\, dx$$

denotes the Fisher information associated to a probability measure μ on $\mathbb{R}$ with differentiable density $d\mu(x) = f(x)\, dx$. Since $J(\nu_1) \ge 1$ in this setting (by the Cramer–Rao inequality), we see that (2.9) always improves upon (2.10), and the improvement can be significant when either $J(\nu_1)$ or $C_p(\nu_1)$ is large. As a particular example, our results apply to uniform measures on convex sets (a prototypical class of measures with finite Poincaré constant), whereas (2.10) degenerates to subaddivity since Fisher information is infinite due to discontinuity of the density at the boundary of its support (in fact, even under arbitrarily small regularization, the Fisher information will still tend to infinity). We additionally note that Fisher information is additive on product measures, so a naïve extension of (2.10) to $\mathbb{R}^d$ would seem to suggest a stability estimate that degrades quickly with dimension.

Johnson's motivation for establishing (2.10) was to quantify convergence of the Poincaré constants $C_p(\nu_n)$ in the CLT, where ν_n is the same as in Example 2.1. In particular, the main result of [19] claims for $d = 1$ and $\sigma^2(\nu_1) = 1$,

$$C_p(\nu_n) \le 1 + \frac{c}{n},$$

where c is a constant depending only on $C_p(\nu_1)$ and $J(\nu_1)$. In actuality, however, this rate of convergence is established for the subsequence (ν_{2^n}), rather than (ν_n) as desired. That is, Theorem 1.2 of [19] should instead state

$$C_p(\nu_n) \le 1 + \frac{c}{\log n}, \tag{2.11}$$

[2]This result is not stated explicitly, but can be distilled from Eq. (3) of [19].

giving an effective rate of convergence of $O(1/\log n)$, rather than $O(n^{-1})$. The mistake appears to be due to a notational oversight in going from the proof of the theorem to the statement of the theorem itself, rather than a technical error.

In view of this, we take the opportunity to revisit the topic of convergence of the Poincaré constant. Unfortunately, despite the improvements of (2.9) over (2.10) described above, our Corollary 2.2 seems incapable of showing that $C_p(\nu_n)-1$ decays asymptotically better than $O(1/\log n)$, whereas a rate of $O(1/n)$ would naturally be conjectured as optimal. Although we suffer from this shortcoming in the asymptotic regime, we can positively show that the Poincaré constant for standardized sums of random vectors converges quickly to a universal constant (i.e., 3/2), before the slower convergence rate kicks in. The precise statement is as follows:

Theorem 2.4. *Let $X_1, X_2, \dots$ be i.i.d. isotropic random vectors in $\mathbb{R}^d$, and define ν_n to be the law of the standardized sum $\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. Then*

$$C_p(\nu_{2^n}) - \frac{3}{2} \le \left(\frac{3}{4}\right)^n \left(C_p(\nu_1) - \frac{3}{2}\right) \tag{2.12}$$

and, if $C_p(\nu_1) \le 2$, it further holds that

$$C_p(\nu_{2^n}) - 1 \le \frac{7}{n+7}. \tag{2.13}$$

Remark 2.1. Using the monotone property $C_p(\nu_{n+1}) \le C_p(\nu_n)$ established in (2.1), the above inequalities taken together give an explicit upper bound on the sequence $C_p(\nu_k)$, even when k is not a power of 2. In particular, if $C_p(\nu_{n_0}) < \infty$ for some n_0, then $\lim_{n\longrightarrow\infty} C_p(\nu_n) = 1$ and convergence takes place monotonically. This statement directly parallels the central limit theorems for entropy and Fisher information.

Remark 2.2. The i.i.d. assumption on the sequence (X_i) may be relaxed to independence with uniformly bounded Poincaré constants.

3. Proofs of main results

3.1. *A variance inequality*

This section is devoted to the proof of a particular variance inequality, from which our main results will follow. First, recall that for a measurable space $\mathcal{X}$, and two probability measures $P \ll Q$ on $\mathcal{X}$, the relative entropy between P and Q is defined as

$$D(P\|Q) := \int_{\mathcal{X}} \log\left(\frac{dP}{dQ}\right) dP.$$

Our starting point is a well-known projection-type inequality enjoyed by relative entropy known as Shearer's lemma (finding origins in [8]), which generalizes inequalities due to Han [17]. Before stating the result, we establish some notation. Let $(\mathcal{X}_i, d_i)$, $i = 1, \dots, n$, be a collection of separable complete metric spaces. If P is a probability measure on the product space $\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$, let P_S denote the corresponding marginal distribution on $\mathcal{X}_S := \prod_{i\in S}\mathcal{X}_i$, where $S \subset \{1, \dots, n\}$. That is, $P_S = \pi_S \# P$, where $\pi_S : \mathcal{X} \longrightarrow \mathcal{X}_S$ is the natural projection. With this notation, Shearer's lemma is the following:

Theorem 3.1. *Let P, Q be Borel probability measures on $\mathcal{X}$, where Q has product form $Q = \prod_{i=1}^n Q_i$. For any collection $\mathcal{C}$ of distinct subsets of $\{1, \dots, n\}$,*

$$\sum_{S\in\mathcal{C}} D(P_S\|Q_S) \le r D(P\|Q), \tag{3.1}$$

where $r := \max_i \#\{S \in \mathcal{C} : S \ni i\}$.

The key inequality we shall need in the present paper is the following corollary, which is obtained by linearizing (3.1):

Corollary 3.1. *With notation as above, let Q be a probability measure on $\mathcal{X}$ with product form, and for the random vector $X=(X_1,\dots,X_n)\sim Q$, let $X_S=(X_i)_{i\in S}$ be the natural projection of X onto $\mathcal{X}_S$. For any collection $\mathcal{C}$ of distinct subsets of $\{1,\dots,n\}$, and any $f:\mathcal{X}\longrightarrow\mathbb{R}$,*

$$\sum_{S\in\mathcal{C}}\operatorname{Var}\bigl(\mathbb{E}\bigl[f(X)|X_S\bigr]\bigr)\le r\operatorname{Var}\bigl(f(X)\bigr), \tag{3.2}$$

where the (conditional) expectation is with respect to Q and $r:=\max_{i\in[n]}\#\{S\in\mathcal{C}:S\ni i\}$.

For completeness, we first prove Theorem 3.1, and then give the derivation of (3.2). We remark that the assumption that $\mathcal{X}$ is a product of separable complete metric spaces ensures that the disintegration theorem can be applied to the measure P in (3.1), which is needed for the proof. This assumption is more than sufficient for our purposes, where we consider only the case where $\mathcal{X}_i=\mathbb{R}^d$.

Proof of Theorem 3.1. For an integer $k\ge 1$, define $[k]=\{1,\dots,k\}$, and $S^k=S\cap[k]$ for $S\subset[n]$. Further, for $S,T\subset\{1,\dots,n\}$, let $P_{S|T}(\cdot|x)$ denote the conditional distribution of P on $\mathcal{X}_S$, given $x\in\mathcal{X}_T$. With notation established, the proof is a simple consequence of properties of relative entropy (cf. [10]):

$$\begin{aligned}
\sum_{S\in\mathcal{C}}D(P_S\|Q_S)&=\sum_{S\in\mathcal{C}}\sum_{i\in S}\int_{\mathcal{X}_{S^{i-1}}}D\bigl(P_{i|S^{i-1}}(\cdot|x)\|Q_i\bigr)\,dP_{S^{i-1}}(x) && (3.3)\\
&\le\sum_{S\in\mathcal{C}}\sum_{i\in S}\int_{\mathcal{X}_{[i-1]}}D\bigl(P_{i|[i-1]}(\cdot|x)\|Q_i\bigr)dP_{[i-1]}(x) && (3.4)\\
&=\sum_{i=1}^{n}\sum_{S\in\mathcal{C}:S\ni i}\int_{\mathcal{X}_{[i-1]}}D\bigl(P_{i|[i-1]}(\cdot|x)\|Q_i\bigr)\,dP_{[i-1]}(x)\\
&\le r\sum_{i=1}^{n}\int_{\mathcal{X}_{[i-1]}}D\bigl(P_{i|[i-1]}(\cdot|x)\|Q_i\bigr)\,dP_{[i-1]}(x) && (3.5)\\
&=r\,D(P\|Q). && (3.6)
\end{aligned}$$

In the above, (3.3) is the chain-rule decomposition of $D(P_S\|Q_S)$; (3.4) is due to convexity of D; (3.5) follows from the definition of r; and (3.6) is the chain-rule decomposition of $D(P\|Q)$. □

Proof of Corollary 3.1. The proof follows by linearizing (3.1), similar to the approach Rothaus used to derive a Poincaré inequality from a logarithmic Sobolev inequality [23]. In particular, we may assume $f:\mathcal{X}\longrightarrow\mathbb{R}$ is bounded with $\mathbb{E}f(X)=\int f\,dQ=0$; the general claim follows by density. Now, for sufficiently small ϵ, define the probability measure P via $dP=(1+\epsilon f)\,dQ$ and observe that Taylor expansion of $D(P\|Q)$ about $\epsilon=0$ gives

$$D\bigl((1+\epsilon f)Q\|Q\bigr)=\frac{\epsilon^2}{2}\int f^2dQ+o\bigl(\epsilon^2\bigr)=\frac{\epsilon^2}{2}\operatorname{Var}\bigl(f(X)\bigr)+o\bigl(\epsilon^2\bigr).$$

More generally, in the notation of Theorem 3.1, we have $dP_S(x_s)=(1+\epsilon\mathbb{E}[f(X)|X_S=x_S])\,dQ_S(x_S)$, so that

$$D(P_S\|Q_S)=\frac{\epsilon^2}{2}\int\bigl(\mathbb{E}\bigl[f(X)|X_S=x_S\bigr]\bigr)^2dQ_S(x_S)+o\bigl(\epsilon^2\bigr)=\frac{\epsilon^2}{2}\operatorname{Var}\bigl(\mathbb{E}\bigl[f(X)|X_S\bigr]\bigr)+o\bigl(\epsilon^2\bigr)$$

for any subset $S\subset[n]$. Hence, substitution into (3.1) and letting ϵ vanish implies (3.2). □

Despite the fundamental nature of (3.2), we could not find any explicit appearance of it in the literature, though it may already be known to some readers. In fact, as pointed out by Y. Polyanskiy, Shearer's lemma holds for any non-negative submodular set function (however, the easiest way to verify the hypothesis for the set function $S\longmapsto\operatorname{Var}(\mathbb{E}[f(X)|X_S])$ may be through a linearization argument applied to entropy, as above). In any case, we are aware of a few related results, which we now briefly discuss.

Remark 3.1. A simple modification gives the following: Let $X_1,\dots,X_n$ be mutually independent random vectors in $\mathbb{R}^d$, and define the random sums $U_S=\sum_{i\in S}X_i$. In this case, for any collection $\mathcal{C}$ of distinct subsets of $\{1,\dots,n\}$ and

$f:\mathbb{R}^d \longrightarrow \mathbb{R}$,

$$\sum_{S\in\mathcal{C}} \mathrm{Var}\big(\mathbb{E}\big[f(U_{[n]})|U_S\big]\big) \le r\,\mathrm{Var}\big(f(U_{[n]})\big),$$

where, as before, $r := \max_{i\in[n]} \#\{S\in\mathcal{C} : S \ni i\}$. In particular, if (X_i) are i.i.d., then a simple consequence is the inequality

$$\mathrm{Var}\big(\mathbb{E}\big[f(U_{[n]})|U_{[m]}\big]\big) \le \frac{m}{n}\mathrm{Var}\big(f(U_{[n]})\big), \quad 1\le m\le n,$$

which is equivalent to the main result of Dembo, Kagan and Shepp in [11].

Remark 3.2. As mentioned in the remarks following Example 2.1, monotonicity of the Poincaré constants of the sequence $(\nu_n)_{n\ge 1}$ parallels the same property enjoyed by entropy and Fisher information, first proved in [1]. In fact, the subset inequality of Theorem 2.1 reminds one of similar subset inequalities enjoyed by entropy and Fisher information, which were proved by Madiman and Barron [22]. This relationship is not coincidental, and we explain the connection here. Specifically, the critical estimate needed in [22] is a "variance drop" inequality of the form

$$\mathbb{E}\bigg|\sum_{S\in\mathcal{C}}\psi_S(X_S)\bigg|^2 \le r\sum_{S\in\mathcal{C}}\mathbb{E}\big|\psi_S(X_S)\big|^2, \tag{3.7}$$

where the notation X_S is the same as above for independent $(X_i)_{i\in[n]}$, and $\psi_S : \mathcal{X}_S \longrightarrow \mathbb{R}$, $S\in\mathcal{C}$, are any functions satisfying $\mathbb{E}\psi_S(X_S)=0$ for each $S\in\mathcal{C}$. Madiman and Barron proved (3.7) using ANOVA decompositions, and apply it to monotonicity of entropy/Fisher information by setting $(\psi_S)_{S\in\mathcal{C}}$ to be score functions of partial sums. As they noted in their paper, (3.7) generalizes a classical result on U-statistics due to Hoeffding [18].

Since (3.2) plays a central role in the proof of Theorem 2.1, the connection between monotonicity of Poincaré constants and Fisher information/entropy in the CLT can be realized through the connection between (3.2) and (3.7). In particular, a new proof of (3.7) can be obtained from (3.2) as follows: Identifying $f(x) := \sum_{S\in\mathcal{C}}\psi_S(x_S)$ and applying Cauchy–Schwarz twice followed by (3.2), we have

$$\begin{aligned}\mathbb{E}\bigg|\sum_{S\in\mathcal{C}}\psi_S(X_S)\bigg|^2 &= \frac{(\sum_{S\in\mathcal{C}}\mathbb{E}f(X)\psi_S(X_S))^2}{\mathbb{E}|f(X)|^2}\\ &\le \frac{(\sum_{S\in\mathcal{C}}(\mathbb{E}|\mathbb{E}[f(X)|X_S]|^2)^{1/2}(\mathbb{E}|\psi_S(X_S)|^2)^{1/2})^2}{\mathbb{E}|f(X)|^2}\\ &\le \bigg(\frac{\sum_{S\in\mathcal{C}}\mathbb{E}|\mathbb{E}[f(X)|X_S]|^2}{\mathbb{E}|f(X)|^2}\bigg)\bigg(\sum_{S\in\mathcal{C}}\mathbb{E}\big|\psi_S(X_S)\big|^2\bigg)\\ &\le r\sum_{S\in\mathcal{C}}\mathbb{E}\big|\psi_S(X_S)\big|^2.\end{aligned}$$

Hence, the monotonicity results for entropy, Fisher information and the Poincaré constant are seen to have a common root in Shearer's inequality.

With (3.2) in hand, the proof of Theorem 2.1 now follows readily.

Proof of Theorem 2.1. For the given collection $\mathcal{C}$, let us define $\bar{\mathcal{C}}$ according to $T\in\bar{\mathcal{C}} \Leftrightarrow ([n]\setminus T)\in\mathcal{C}$. For independent random vectors $X_i\sim\mu_i$, $i\in[n]$, and $T\subset[n]$, define $X_T=(X_i)_{i\in T}$, and let $U=\sum_{i=1}^n X_i$. Consider any smooth $f:\mathbb{R}^d\longrightarrow\mathbb{R}$. For any $T\subset[n]$, we have the classical variance decomposition

$$\mathrm{Var}\big(f(U)\big) = \mathbb{E}\big[\mathrm{Var}\big(f(U)|X_T\big)\big] + \mathrm{Var}\big(\mathbb{E}\big[f(U)|X_T\big]\big).$$

Summing over subsets $T\in\bar{\mathcal{C}}$ and applying (3.2), we find

$$\begin{aligned}|\bar{\mathcal{C}}|\,\mathrm{Var}\big(f(U)\big) &= \sum_{T\in\bar{\mathcal{C}}}\mathbb{E}\big[\mathrm{Var}\big(f(U)|X_T\big)\big] + \sum_{T\in\bar{\mathcal{C}}}\mathrm{Var}\big(\mathbb{E}\big[f(U)|X_T\big]\big)\\ &\le \sum_{T\in\bar{\mathcal{C}}}\mathbb{E}\big[\mathrm{Var}\big(f(U)|X_T\big)\big] + r\,\mathrm{Var}\big(f(U)\big),\end{aligned}$$

where $r := \max_{i\in[n]} \#\{T \in \bar{\mathcal{C}} : T \ni i\}$. Rearranging and applying the Poincaré inequality for $\mu_{[n]\setminus T}$, $T \in \bar{\mathcal{C}}$, we have

$$\begin{aligned}\big(|\bar{\mathcal{C}}| - r\big)\operatorname{Var}\big(f(U)\big) &\le \sum_{T\in\bar{\mathcal{C}}} \mathbb{E}\big[\operatorname{Var}\big(f(U)|X_T\big)\big] \\ &\le \sum_{T\in\bar{\mathcal{C}}} \mathbb{E}\big[C_p(\mu_{[n]\setminus T})\mathbb{E}\big[\big|\nabla f(U)\big|^2|X_T\big]\big] \\ &= \sum_{S\in\mathcal{C}} C_p(\mu_S)\mathbb{E}\big[\big|\nabla f(U)\big|^2\big].\end{aligned}$$

Now, the proof is complete by noting that $(|\bar{\mathcal{C}}| - r) = \min_{i\in[n]} \#\{S \in \mathcal{C} : S \ni i\}$. □

Remark 3.3. The proof of Theorem 2.2 is identical, and is therefore omitted.

3.2. *Proof of Theorems* 2.3 *and* 2.4

This section is dedicated to the proofs of Theorems 2.3 and 2.4. We begin with a technical lemma. First, a remark on notation throughout this section:

Remark 3.4. For a vector-valued function $g : \mathcal{X} \longrightarrow \mathbb{R}^d$ and a probability measure μ on $\mathcal{X}$, we abuse notation slightly and write $\operatorname{Var}_\mu(g)$ to denote $\sum_{i=1}^d \operatorname{Var}_\mu(g_i)$, where $g_i : \mathcal{X} \longrightarrow \mathbb{R}$ denotes the ith coordinate of $g = (g_1, \dots, g_d)$. In particular,

$$\operatorname{Var}_\mu(g) := \int |g|^2\, d\mu - \left|\int g\, d\mu\right|^2.$$

Lemma 3.1. *Let μ be a probability measure on $\mathbb{R}^d$ which verifies a Poincaré inequality with best constant $C_p = C_p(\mu)$. Define $\sigma^2 := \sup_{\alpha:|\alpha|=1} \operatorname{Var}_\mu(x \mapsto \alpha\cdot x)$. There is a sequence (f_n) of real-valued functions on $\mathbb{R}^d$ with $\int |\nabla f_n|^2\, d\mu = 1$, which satisfies*

$$\lim_{n\longrightarrow\infty} \operatorname{Var}_\mu(f_n) = C_p,$$

and

$$\lim_{n\longrightarrow\infty} \operatorname{Var}_\mu(\nabla f_n) \ge \frac{(C_p - \sigma^2)^2}{(C_p - \sigma^2)^2 + C_p\sigma^2}.$$

Remark 3.5. The idea here is that if $C_p > \sigma^2$, then there are near-extremizers of the Poincaré inequality for μ which have nontrivial projection onto the space of nonlinear functions. As a result, the variances $\operatorname{Var}_\mu(\nabla f_n)$ can be nontrivially compared to the moments $\int |\nabla f_n|^2\, d\mu$. This result is suggested by Johnson for dimension 1 in [19], but is only formally argued under the assumption that an extremizer exists for the Poincaré inequality. The potential nonexistence of extremizers is the main issue to be dealt with, and can be handled through an application of the Lax–Milgram theorem, as below.

Proof. We first show that if f is sufficiently smooth, satisfying $\int f\, d\mu = 0$ and

$$\int |f|^2\, d\mu \ge \big(1 - \epsilon^2\big)C_p \int |\nabla f|^2\, d\mu, \tag{3.8}$$

then

$$C_p \int \nabla f \cdot \nabla h\, d\mu - \int fh\, d\mu \le C_p\epsilon\left(\int |\nabla f|^2\, d\mu\right)^{1/2}\left(\int |\nabla h|^2\, d\mu\right)^{1/2} \tag{3.9}$$

for all sufficiently smooth h. This may be seen as a stable form of the Euler–Lagrange equation associated to the Poincaré inequality for μ. Indeed, if there exists a nonzero function f_0 which $\operatorname{Var}_\mu(f_0) = C_p \int |\nabla f_0|^2\, d\mu$, then

$$C_p \int \nabla f_0 \cdot \nabla h\, d\mu = \int f_0 h\, d\mu$$

for all sufficiently smooth h.

Toward establishing (3.9), consider the Sobolev space $W^{1,2}$, defined as the closure of the set of functions $\{f \in C^\infty(\mathbb{R}^d) : \int f\,d\mu = 0\}$ in $L^2(\mu)$ with respect to the Sobolev norm $\|f\| := (\int |\nabla f|^2\,d\mu + \int |f|^2\,d\mu)^{1/2}$.

By the Poincaré inequality, the continuous bilinear map $(f,g) \mapsto C_p \int \nabla f \cdot \nabla g\,d\mu$ is coercive on $W^{1,2} \times W^{1,2}$. So, for any $f \in W^{1,2}$, the Lax–Milgram theorem ensures the existence of $u_f \in W^{1,2}$ such that

$$C_p \int \nabla u_f \cdot \nabla h\,d\mu = \int f h\,d\mu \quad \text{for all } h \in W^{1,2}.$$

Applying this to the function $h = u_f$, Cauchy–Schwarz gives

$$C_p \int |\nabla u_f|^2\,d\mu = \int u_f f\,d\mu \le \left(C_p \int |\nabla u_f|^2\,d\mu\right)^{1/2} \left(\int |f|^2\,d\mu\right)^{1/2},$$

so that $C_p \int |\nabla u_f|^2\,d\mu \le \int |f|^2\,d\mu$. Now, if f verifies (3.8), then

$$\begin{aligned} C_p \int |\nabla u_f - \nabla f|^2\,d\mu &= C_p \int |\nabla u_f|^2\,d\mu + C_p \int |\nabla f|^2\,d\mu - 2C_p \int \nabla u_f \cdot \nabla f\,d\mu \\ &= C_p \int |\nabla u_f|^2\,d\mu + C_p \int |\nabla f|^2\,d\mu - 2\int |f|^2\,d\mu \\ &\le \epsilon^2 C_p \int |\nabla f|^2\,d\mu. \end{aligned}$$

As a consequence, we obtain

$$\begin{aligned} C_p \int \nabla f \cdot \nabla h\,d\mu - \int f h\,d\mu &= C_p \int (\nabla f - \nabla u_f) \cdot \nabla h\,d\mu \\ &\le C_p \epsilon \left(\int |\nabla f|^2\,d\mu\right)^{1/2} \left(\int |\nabla h|^2\,d\mu\right)^{1/2}, \end{aligned}$$

which is (3.9).

Henceforth, we assume f satisfies $\int f\,d\mu = 0$ and (3.8). We may also assume without loss of generality that $\int x\,d\mu(x) = 0$, since translation does not change the Poincaré constant. For any $\alpha \in \mathbb{R}^d$, definition of σ^2 together with the Poincaré and Cauchy–Schwarz inequalities yields

$$\begin{aligned} \int f(x)(\alpha \cdot x)\,d\mu(x) - |\alpha|^2\sigma^2 &= \int \big(f(x) - \alpha \cdot x\big)(\alpha \cdot x)\,d\mu(x) \\ &\le \left(\int \big|f(x) - \alpha \cdot x\big|^2\,d\mu(x)\right)^{1/2} \left(\int |\alpha \cdot x|^2\,d\mu(x)\right)^{1/2} \\ &\le |\alpha|\sigma \left(C_p \int |\nabla f - \alpha|^2\,d\mu\right)^{1/2}. \end{aligned}$$

Applying (3.9) with $h(x) = \alpha \cdot x$, we conclude

$$C_p \int \nabla f \cdot \alpha\,d\mu - |\alpha|^2\sigma^2 \le |\alpha| \left(C_p \int |\nabla f - \alpha|^2\,d\mu\right)^{1/2} \sigma + C_p \epsilon \left(\int |\nabla f|^2\,d\mu\right)^{1/2} |\alpha|.$$

Specializing this by taking $\alpha = \int \nabla f\,d\mu$, we find

$$\begin{aligned} &\big(C_p - \sigma^2\big)\left(\int |\nabla f|^2\,d\mu - \mathrm{Var}_\mu(\nabla f)\right) \\ &\quad \le \left(\int |\nabla f|^2\,d\mu - \mathrm{Var}_\mu(\nabla f)\right)^{1/2} \left(\big(C_p \,\mathrm{Var}_\mu(\nabla f)\big)^{1/2}\sigma + C_p \epsilon \left(\int |\nabla f|^2\,d\mu\right)^{1/2}\right). \end{aligned} \tag{3.10}$$

At this point, the proof is essentially complete. Indeed, by homogeneity of the Poincaré inequality, we can find a sequence $(f_n) \subset W^{1,2}$ such that $\int |\nabla f_n|^2\,d\mu = 1$ and

$$\lim_{n \longrightarrow \infty} \mathrm{Var}_\mu(f_n) = C_p.$$

Since $\mathrm{Var}_\mu(\nabla f_n) \le \int |\nabla f_n|^2\, d\mu = 1$, we may extract a subsequence for which the limit $\lim_{k\longrightarrow\infty} \mathrm{Var}_\mu(\nabla f_{n_k}) =: V_\infty$ exists. Applying (3.10) to this subsequence, we may let $\epsilon \downarrow 0$ to conclude

$$\big(C_p - \sigma^2\big)\sqrt{1 - V_\infty} \le \sigma\sqrt{C_p}\sqrt{V_\infty}. \tag{3.11}$$

Squaring both sides and rearranging completes the proof. □

Now, with the help of Lemma 3.1 and (3.2), we are ready to prove Theorem 2.3.

Proof of Theorem 2.3. For convenience, we use probabilistic notation with $X \sim \mu$ and $Y \sim \nu$ independent, and $U = X + Y$. With this notation, we first aim to show that

$$\mathrm{Var}\big(f(U)\big) \le \big(C_p(\mu) + C_p(\nu)\big)\mathbb{E}\big[\big|\nabla f(U)\big|^2\big] - \frac{C_p(\mu)C_p(\nu)}{C_p(\mu) + C_p(\nu)}\mathrm{Var}\big(\nabla f(U)\big) \tag{3.12}$$

for differentiable f. To this end, we consider a smooth test function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$; the general result follows from density. Without loss of generality, we may also assume $\mathbb{E}[f(U)] = 0$. We have the classical variance decomposition

$$\begin{aligned}\mathrm{Var}\big(f(U)\big) &= \mathbb{E}\,\mathrm{Var}\big(f(U)|X\big) + \mathrm{Var}\big(\mathbb{E}\big[f(U)|X\big]\big)\\ &=: A + B.\end{aligned}$$

As in the proof of Theorem 2.1, since ν satisfies a Poincaré inequality with constant $C_p(\nu)$, the first term A is bounded by

$$A = \mathbb{E}\,\mathrm{Var}\big(f(U)|X\big) \le \mathbb{E}\big[C_p(\nu)\mathbb{E}\big[\big|\nabla f(U)\big|^2|X\big]\big] = C_p(\nu)\mathbb{E}\big[\big|\nabla f(U)\big|^2\big].$$

Departing from the proof of Theorem 2.1, we bound the second term B as

$$B = \mathrm{Var}\big(\mathbb{E}\big[f(U)|X\big]\big) \le C_p(\mu)\mathbb{E}\big|\nabla\mathbb{E}\big[f(U)|X\big]\big|^2 = C_p(\mu)\mathbb{E}\big|\mathbb{E}\big[\nabla f(U)|X\big]\big|^2,$$

where moving the gradient inside the expectation is justified by smoothness of f. Written another way, we have

$$B \le C_p(\mu)\big(\mathrm{Var}\big(\mathbb{E}\big[\nabla f(U)|X\big]\big) + \big|\mathbb{E}\big[\nabla f(U)\big]\big|^2\big).$$

By symmetry, we obtain a similar bound with the roles of μ and ν (resp. X and Y) reversed. Hence, taking a convex combination of these two separate bounds, we find

$$\begin{aligned}\mathrm{Var}\big(f(U)\big) \le{}& \frac{C_p(\nu)^2 + C_p(\mu)^2}{C_p(\mu) + C_p(\nu)}\mathbb{E}\big[\big|\nabla f(U)\big|^2\big] + 2\frac{C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\big|\mathbb{E}\big[\nabla f(U)\big]\big|^2\\ &+ \frac{C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\big(\mathrm{Var}\big(\mathbb{E}\big[\nabla f(U)|X\big]\big) + \mathrm{Var}\big(\mathbb{E}\big[\nabla f(U)|Y\big]\big)\big).\end{aligned}$$

Applying (3.2) to the sum of variance terms and using the identity $\mathrm{Var}(\nabla f(U)) = \mathbb{E}[|\nabla f(U)|^2] - |\mathbb{E}[\nabla f(U)]|^2$, we have

$$\begin{aligned}\mathrm{Var}\big(f(U)\big) &\le \frac{C_p(\nu)^2 + C_p(\mu)^2 + C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\mathbb{E}\big[\big|\nabla f(U)\big|^2\big] + \frac{C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\big|\mathbb{E}\big[\nabla f(U)\big]\big|^2\\ &= \frac{C_p(\nu)^2 + C_p(\mu)^2 + 2C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\mathbb{E}\big[\big|\nabla f(U)\big|^2\big] - \frac{C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\mathrm{Var}\big(\nabla f(U)\big)\\ &= \big(C_p(\mu) + C_p(\nu)\big)\mathbb{E}\big[\big|\nabla f(U)\big|^2\big] - \frac{C_p(\nu)C_p(\mu)}{C_p(\mu) + C_p(\nu)}\mathrm{Var}\big(\nabla f(U)\big).\end{aligned}$$

The first equality follows again from the identity $\mathrm{Var}(\nabla f(U)) = \mathbb{E}[|\nabla f(U)|^2] - |\mathbb{E}[\nabla f(U)]|^2$. So, (3.12) is proved.

At this point, we need to deal with the term $\mathrm{Var}(\nabla f(U))$ in (3.12) in order to bound $C_p(\mu \star \nu)$. To this end, Lemma 3.1 ensures the existence of a sequence (f_n) with $\mathbb{E}[|\nabla f_n(U)|^2] = 1$, which satisfies

$$\lim_{n\longrightarrow\infty} \mathrm{Var}\big(f_n(U)\big) = C_p(\mu \star \nu)$$

and

$$\lim_{n\longrightarrow\infty} \mathrm{Var}\big(\nabla f_n(U)\big) \geq \frac{(C_p(\mu\star\nu)-\sigma^2)^2}{(C_p(\mu\star\nu)-\sigma^2)^2+C_p(\mu\star\nu)\sigma^2}, \tag{3.13}$$

where $\sigma^2 := \sigma^2(\mu\star\nu)$ as in the statement of the theorem. Substituting this sequence into (3.12) and bounding the variance terms with (3.13) completes the proof. □

Remark 3.6. The above proof strategy does not appear to easily extend show stability of (2.1). The reason is that the separate bounds on the quantities A and B cause both Poincaré constants $C_p(\mu)$ and $C_p(\nu)$ to appear in bounding $C_p(\mu\star\nu)$.

Remark 3.7. We remark that the above proof closely follows the development in [19]. The key difference is that our use of the variance inequality (3.2) avoids the introduction of Fisher information as in Johnson's proof.

At this point, we need only to prove Theorem 2.4.

Proof of Theorem 2.4. Starting with Corollary 2.2, we have

$$\begin{aligned} C_p(\nu_2) &\leq C_p(\nu_1) - \frac{C_p(\nu_1)}{4}\frac{(C_p(\nu_2)-1)^2}{(C_p(\nu_2)-1)^2+C_p(\nu_2)} \\ &= C_p(\nu_1)\left(1-\frac{1}{4}\frac{(C_p(\nu_2)-1)^2}{(C_p(\nu_2)-1)^2+C_p(\nu_2)}\right) \end{aligned}$$

since $\sigma^2(\nu_2)=1$ due to the isotropic assumption. On rearranging, we find

$$\begin{aligned} C_p(\nu_1)-C_p(\nu_2) &\geq \frac{C_p(\nu_2)(C_p(\nu_2)-1)^2}{3(C_p(\nu_2)-1)^2+4C_p(\nu_2)} \\ &\geq \begin{cases} \frac{1}{6}(C_p(\nu_2)-1)^2 & 1\leq C_p(\nu_2)<2 \\ \frac{1}{3}(C_p(\nu_2)-1)-1/6 & C_p(\nu_2)\geq 1, \end{cases} \end{aligned} \tag{3.14}$$

where the second inequality follows from elementary calculus. Using the affine lower bound in (3.14), a straightforward inductive argument gives

$$C_p(\nu_{2^n})-1 \leq \frac{1}{2}\left(1-\left(\frac{3}{4}\right)^n\right)+\left(\frac{3}{4}\right)^n\big(C_p(\nu_1)-1\big),$$

which is (2.12).

To establish (2.13), construct a sequence $(a_n)_{n\geq 0}$ inductively starting with $a_0=2$, and defining a_{n+1} to be the positive root of the quadratic equation

$$a_{n+1}+\frac{1}{6}(a_{n+1}-1)^2 = a_n, \quad n\geq 0. \tag{3.15}$$

If $1\leq C_p(\nu_1)\leq 2$, the quadratic bound in (3.14) implies

$$C_p(\nu_{2^{n+1}})+\frac{1}{6}\big(C_p(\nu_{2^{n+1}})-1\big)^2 \leq C_p(\nu_{2^n})$$

for $n\geq 0$, so that we necessarily have $C_p(\nu_{2^n})\leq a_n$ for all $n\geq 0$. Hence, we only need to upper bound the sequence $(a_n)_{n\geq 0}$. To that end, applying the quadratic formula to (3.15), we have

$$a_{n+1} = -2+\sqrt{9+6(a_n-1)}, \quad n\geq 0.$$

We claim that $a_n\leq 1+\frac{7}{n+7}$. Indeed, this is true for $n=0$ by definition. By induction,

$$a_{n+1} = -2+\sqrt{9+6(a_n-1)} \leq -2+\sqrt{9+6\frac{7}{n+7}} \leq 1+\frac{7}{(n+1)+7},$$

where the last inequality can be checked to hold for all $n\geq -1$. This completes the proof. □

Acknowledgements

This work was supported by the France-Berkeley fund and NSF grants CCF-1750430 and CCF-0939370. The author thanks Max Fathi, Oliver Johnson and Michel Ledoux for comments and discussions. We also would like to acknowledge the hospitality of the Modern Mathematical Methods for Data Analysis workshop held in Liège Université, and the Workshop on Stability of Functional Inequalities and Applications at Université Paul Sabatier, Institut de Mathématiques de Toulouse.

References

[1] S. Artstein, K. Ball, F. Barthe and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. *J. Amer. Math. Soc.* **17** (4) (2004) 975–982. MR2083473 https://doi.org/10.1090/S0894-0347-04-00459-X

[2] D. Bakry, I. Gentil and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Springer, Cham, 2014. MR3155209 https://doi.org/10.1007/978-3-319-00227-9

[3] J.-B. Bardet, N. Gozlan, F. Malrieu and P.-A. Zitt. Functional inequalities for Gaussian convolutions of compactly supported measures: Explicit bounds and dimension dependence. *Bernoulli* **24** (1) (2018) 333–353. MR3706760 https://doi.org/10.3150/16-BEJ879

[4] S. G. Bobkov, G. P. Chistyakov and F. Götze. Berry–Esseen bounds in the entropic central limit theorem. *Probab. Theory Related Fields* **159** (3–4) (2014) 435–478. MR3230000 https://doi.org/10.1007/s00440-013-0510-3

[5] S. G. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** (3) (1997) 383–400. MR1440138 https://doi.org/10.1007/s004400050090

[6] A. A. Borovkov and S. A. Utev. On an inequality and a related characterization of the normal distribution. *Teor. Veroyatn. Primen.* **28** (2) (1983) 209–218. MR0700206

[7] D. Chafai and F. Malrieu. On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (1) (2010) 72–96. MR2641771 https://doi.org/10.1214/08-AIHP309

[8] F. R. Chung, R. L. Graham, P. Frankl and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Combin. Theory Ser. A* **43** (1) (1986) 23–37. MR0859293 https://doi.org/10.1016/0097-3165(86)90019-1

[9] T. A. Courtade, M. Fathi and A. Pananjady. Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** (2) (2019) 777–790. MR3949953 https://doi.org/10.1214/18-AIHP898

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory*, 2nd edition. John Wiley & Sons, New York, 2012. MR2239987

[11] A. Dembo, A. Kagan and L. A. Shepp. Remarks on the maximum correlation coefficient. *Bernoulli* **7** (2) (2001) 343–350. MR1828509 https://doi.org/10.2307/3318742

[12] L. C. Evans. *Partial Differential Equations*, 2nd edition. *Graduate Studies in Mathematics* **19**. American Mathematical Society, Providence, RI, 2010. MR2597943 https://doi.org/10.1090/gsm/019

[13] T. Gallouët, G. Mijoule and Y. Swan. Regularity of solutions of the Stein equation and rates in the multivariate central limit theorem. ArXiv e-prints, 2018.

[14] N. Gozlan. Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (3) (2010) 708–739. MR2682264 https://doi.org/10.1214/09-AIHP209

[15] N. Gozlan, C. Roberto and P.-M. Samson. From dimension free concentration to the Poincaré inequality. *Calc. Var. Partial Differential Equations* **52** (3–4) (2015) 899–925. MR3311918 https://doi.org/10.1007/s00526-014-0737-6

[16] M. Gromov and V. D. Milman. A topological application of the isoperimetric inequality. *Amer. J. Math.* **105** (4) (1983) 843–854. MR0708367 https://doi.org/10.2307/2374298

[17] T. S. Han. Nonnegative entropy measures of multivariate symmetric correlations. *Inf. Control* **36** (2) (1978) 133–156. MR0464499

[18] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.* **19** (3) (1948) 293–325. MR0026294 https://doi.org/10.1214/aoms/1177730196

[19] O. Johnson. Convergence of the Poincaré constant. *Teor. Veroyatn. Primen.* **48** (2003) 615–620. MR2141356 https://doi.org/10.1137/S0040585X97980622

[20] M. Ledoux. *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, 2001. MR1849347

[21] M. Ledoux, I. Nourdin and G. Peccati. Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25** (1) (2015) 256–306. MR3320893 https://doi.org/10.1007/s00039-015-0312-0

[22] M. Madiman and A. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory* **53** (7) (2007) 2317–2329. MR2319376 https://doi.org/10.1109/TIT.2007.899484

[23] O. S. Rothaus. Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *J. Funct. Anal.* **42** (1) (1981) 102–109. MR0620581 https://doi.org/10.1016/0022-1236(81)90049-5

[24] G. Valiant and P. Valiant. Estimating the unseen: An $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* 685–694, 2011. MR2932019 https://doi.org/10.1145/1993636.1993727

[25] A. Zhai. A high-dimensional CLT in W_2 distance with near optimal convergence rate. *Probab. Theory Related Fields* **170** (3–4) (2018) 821–845. MR3773801 https://doi.org/10.1007/s00440-017-0771-3

[26] D. Zimmermann. Logarithmic Sobolev inequalities for mollified compactly supported measures. *J. Funct. Anal.* **265** (6) (2013) 1064–1083. MR3067796 https://doi.org/10.1016/j.jfa.2013.05.029